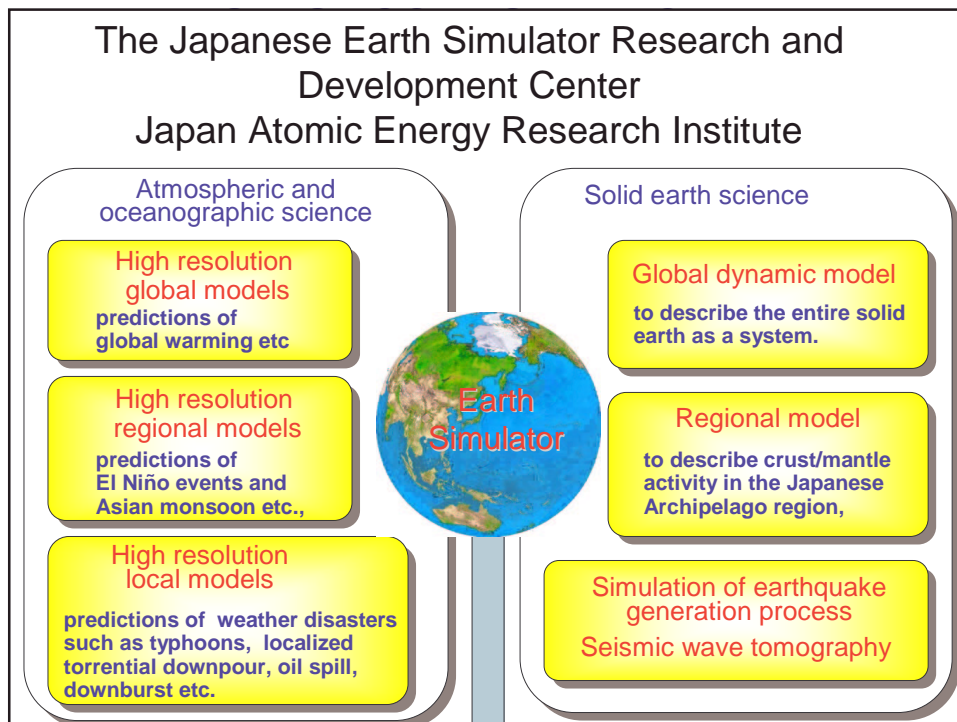


# Notes on the Earth Simulator

Jack Dongarra  
Computer Science Department  
University of Tennessee

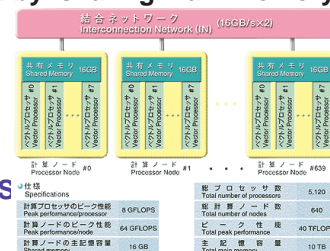


# Earth Simulator

- Based on the NEC SX architecture, 640 nodes, each node with 8 vector processors (8 Gflop/s peak per processor), 2 ns cycle time, 16GB shared memory.
  - Total of 5104 total processors, 40 TFlop/s peak, and 10 TB memory.
- It has a single stage crossbar (1800 miles of cable) 83,000 copper cables, 16 GB/s cross section bandwidth.
- 700 TB disk space
- 1.6 PB mass store
- Area of computer = 4 tennis courts, 3 floors

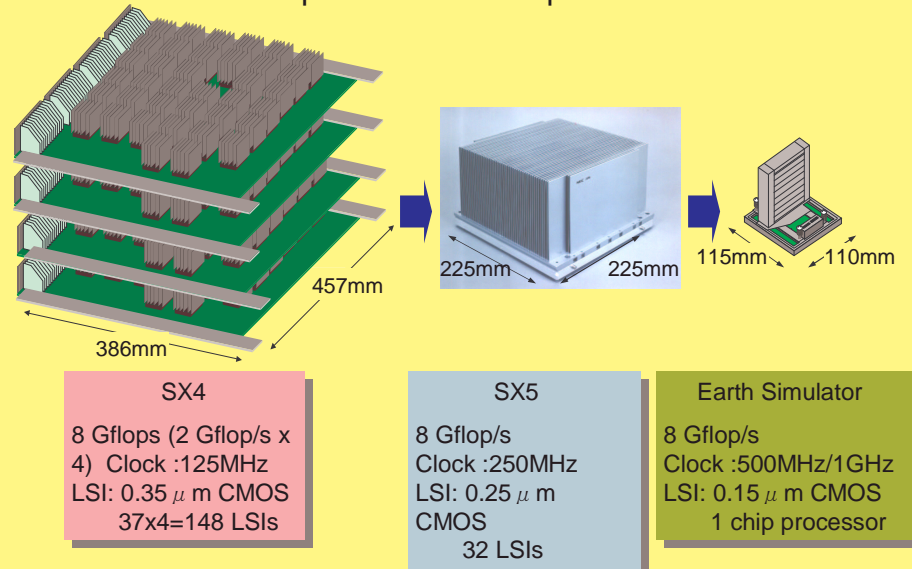
## Outline of the Earth Simulator Computer

- **Architecture** : A MIMD-type, distributed memory, parallel system consisting of computing nodes in which vector-type multi-processors are tightly connected by sharing main memory
- **Total number of processor nodes: 640**
- **Number of PE's for each node: 8**
- **Total number of PE's: 5120**
- **Peak performance of each PE: 8 GFLOPS**
- **Peak performance of each node: 64 GFLOPS**
- **Main memory** : 10 TB (total).  
 Shared memory / node : 16 GB
- **Interconnection network: Single-Stage Crossbar Network**
- **Performance** : Assuming the efficiency 12.5%, the peak performance 40 TFLOPS (the effective performance for an atmospheric circulation model is more than 5 TFLOPS).



## R&D results

### Comparison of vector processors



Earth Simulator Research and Development Center

## R&D results

### R&D Issues on Hardware Technologies

#### (1) LSI Technology

- Enhancement of clock cycle 150MHz  $\Rightarrow$  500MHz (partly 1GHz)
- Development of high density LSI
  - 0.15  $\mu$  m CMOS + Cu interconnection (8 layers)
  - 1.50-2.0 million transistors/cm<sup>2</sup>  $\Rightarrow$  10 million transistors/cm<sup>2</sup>
- Enlargement of chip size (about 2cm x 2cm)

#### High performance one-chip vector processor: OCVP-ES

#### (2) Packaging Technology

- Build-up PCB (110mm x 115mm)
  - Line width / Spacing : 25  $\mu$  m / 25  $\mu$  m
  - 6 core layers + 4 build-up layers on both surfaces
- number of pins/chip <1000 (present)  $\Rightarrow$  4000 - 5000

#### (3) Cooling Technology

- Air cooling using heat pipe technology (Max. 170W per chip)

#### (4) Board to Board Interconnection Technology

- Interface connector 0.5mm pitch surface mount
- Interface cable 0.6mm diameter coaxial cable and 3.8ns/m delay time

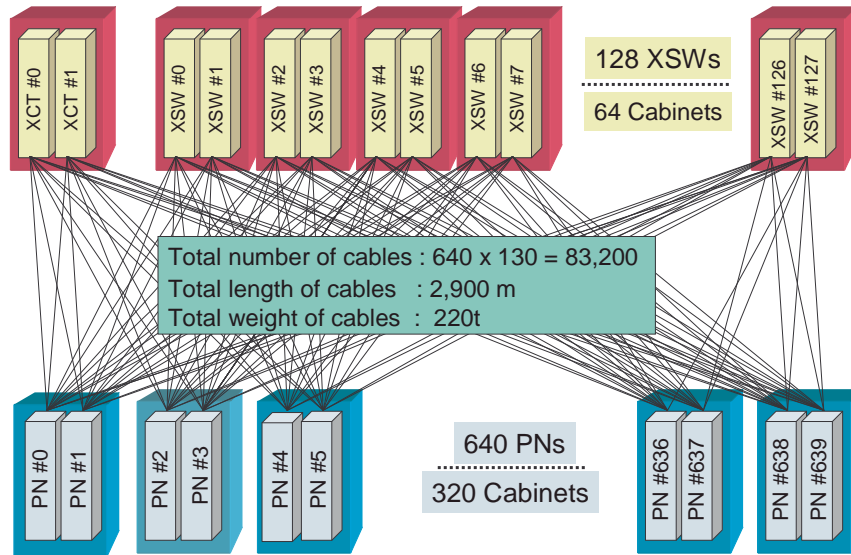
#### (5) PN-IN Interconnection Technology

- 40m transmission distance with fine tuned equalizer circuit

Earth Simulator Research and Development Center

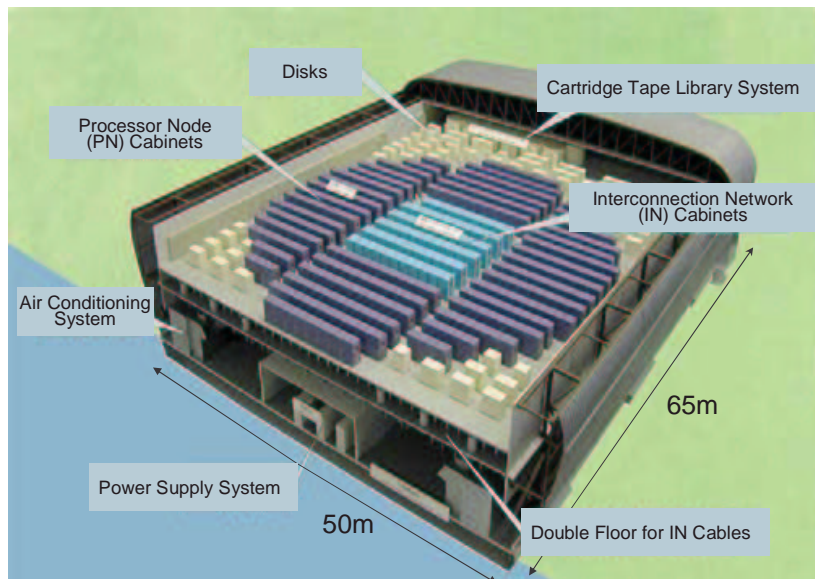
## R&D results

### Connection between processor nodes (crossbar network)

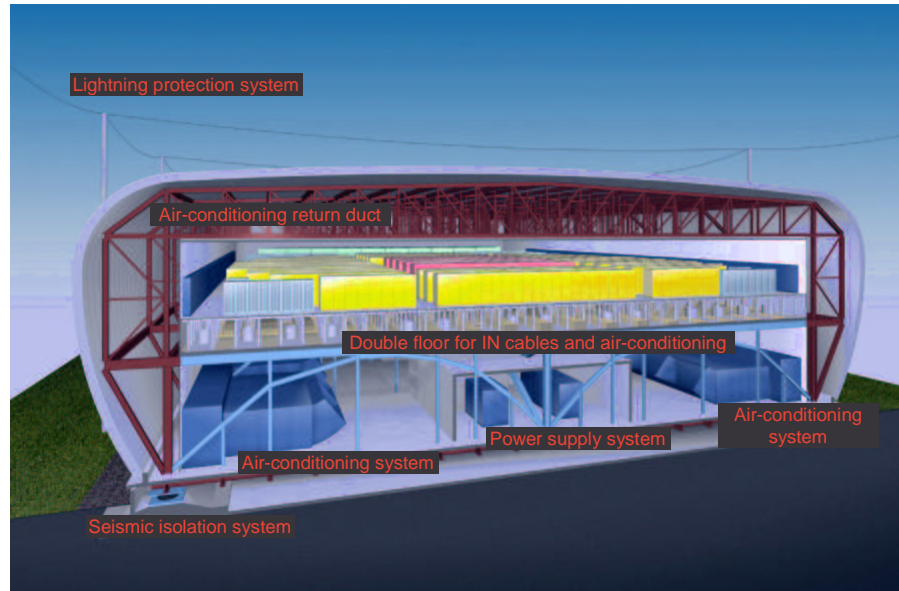


Earth Simulator Research and Development Center

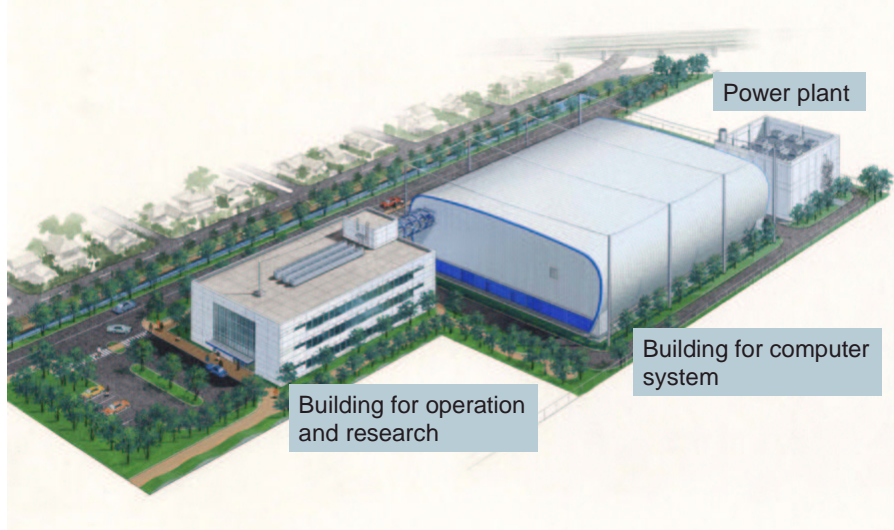
### Bird's-eye View of the Earth Simulator System



Cross-sectional View of the Earth Simulator Building



New Earth Simulator Facilities





### Wiring of interconnection network cables

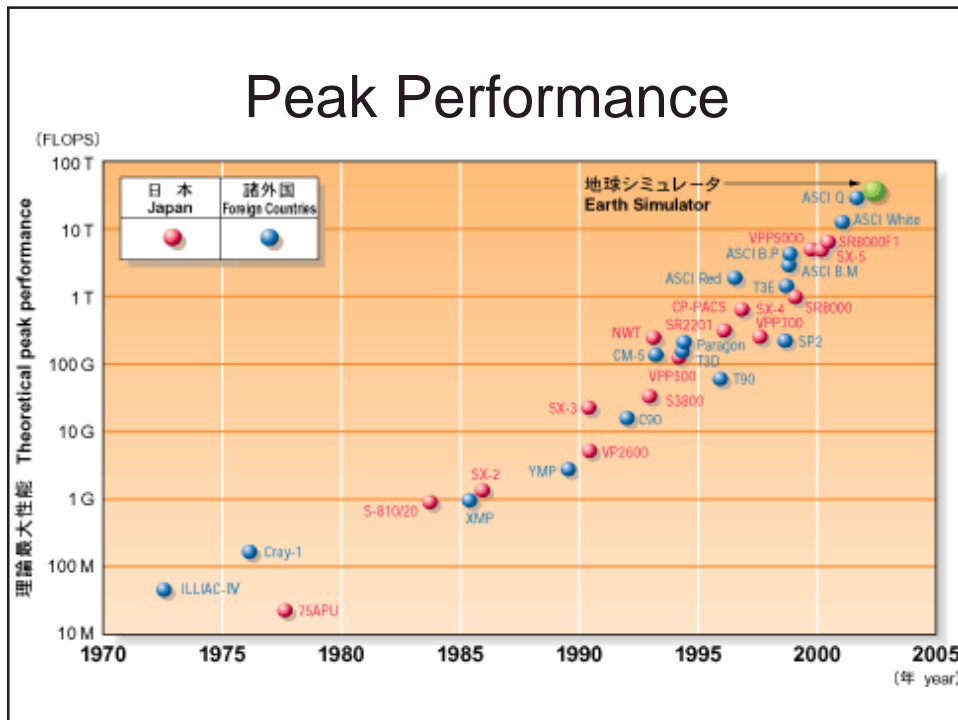


Earth Simulator Research and Development Center

### Wiring of interconnection network cables

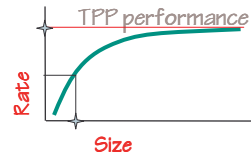


Earth Simulator Research and Development Center



# Earth Simulator Computer (ESC)

- Rmax from LINPACK MPP Benchmark *Ax=b, dense problem*
  - Linpack Benchmark = 35.6 TFlop/s
  - Problem of size  $n = 1,041,216$ ; (8.7 TB of memory)
  - Half of peak ( $n_{1/2}$ ) achieved at  $n_{1/2} = 265,408$
  - Benchmark took 5.8 hours to run.
  - Algorithm: LU w/partial pivoting
  - Software: for the most part Fortran using MPI



- For the Top500
  - $\Sigma$  of all the DOE computers = 24 TFlop/s
  - Performance of ESC  $\sim \frac{1}{4} \Sigma(\text{Top 500 Computers})$
  - Performance of ESC  $> \Sigma(\text{Top 18 Computers})$
  - Performance of ESC  $> \Sigma(\text{Top 20 Computers in the US})$
  - Performance of ESC  $>$  All the DOE and DOD machines (27.6 TFlop/s)
  - Performance of ESC  $\gg$  the 3 NSF Center's computers (8.4 TFlop/s)

SETI@home  $\sim 27$  TFlop/s

## Machine at the Top of the List

Year	Computer	Measured Gflop/s	Factor $\Delta$ from Pervious Year	Theoretical Peak Gflop/s	Factor $\Delta$ from Pervious Year	Number of Processors	Size of Problem
2002	Earth Simulator Computer, NEC	35610	4.9	40832	3.7	5104	1041216
2001	ASCI White-Pacific, IBM SP Power 3	7226	1.5	11136	1.0	7424	518096
2000	ASCI White-Pacific, IBM SP Power 3	4938	2.1	11136	3.5	7424	430000
1999	ASCI Red Intel Pentium II Xeon core	2379	1.1	3207	0.8	9632	362880
1998	ASCI Blue-Pacific SST, IBM SP 604E	2144	1.6	3868	2.1	5808	431344
1997	Intel ASCI Option Red (200 MHz Pentium Pro)	1338	3.6	1830	3.0	9152	235000
1996	Hitachi CP-PACS	368.2	1.3	614	1.8	2048	103680
1995	Intel Paragon XP/S MP	281.1	1	338	1.0	6768	128600
1994	Intel Paragon XP/S MP	281.1	2.3	338	1.4	6768	128600
1993	Fujitsu NWT	124.5		236		140	31920



# LINPACK Benchmark List

Computer (Full Precision)		Number of Processors	$R_{max}$ Gflop/s	$N_{max}$ order	$N_{1/2}$ order	$R_{peak}$ Gflop/s
★Earth Simulator, NEC processors****	esc	5104	35610	1041216	265408	40832
ASCI White-Pacific, IBM SP Power 3(375 MHz)	llnl	8000	7226	518096	179000	12000
★Compaq AlphaServer SC ES45/EV68 1GHz	psc	3016	4463	280000	85000	6032
Compaq AlphaServer SC ES45/EV68 1GHz	psc	3024	4059	525000	105000	6048
★Compaq AlphaServer SC ES45/EV68 1GHz	cea	2560	3980	360000	85000	5120
IBM SP Power3 208 nodes 375 MHz	lbl	3328	3052.	371712		4992
★Compaq Alphaserer SC ES45/EV68 1GHz	lanl	2048	2916	272000		4096
★IBM SP Power3 158 nodes 375 MHz	lbl	2528	2526.	371712	102400	3792
ASCI Red Intel Pentium II Xeon core 333MHz	snl	9632	2379.6	362880	75400	3207
ASCI Blue-Pacific SST, IBM SP 604E(332 MHz)	llnl	5808	2144.	431344	432344	3868
ASCI Red Intel Pentium II Xeon core 333MHz	snl	9472	2121.3	251904	66000	3154
Compaq Alphaserer SC ES45/EV68 1GHz	lanl	1520	2096	390000	71000	3040
★IBM SP 112 nodes (375 MHz POWER3 High)	ibm	1792	1791	275000	275000	2688
HITACHI SR8000/MPP/1152(450MHz)	u toyko	1152	1709.1	141000	16000	2074
★HITACHI SR8000-F1/168(375MHz)	leibniz	168	1653.	160000	19560	2016
ASCI Red Intel Pentium II Xeon core 333Mhz	snl	6720	1633.3	306720	52500	2238
SGI ASCI Blue Mountain	lanl	5040	1608.	374400	138000	2520
IBM SP 328 nodes (375 MHz POWER3 Thin)	noo	1312	1417.	374000	374000	1968
Intel ASCI Option Red (200 MHz Pentium Pro)	snl	9152	1338.	235000	63000	1830
NEC SX-5/128M8(3.2ns)	osaka	128	1192.0	129536	10240	1280
CRAY T3E-1200 (600 MHz)	us government	1488	1127.	148800	28272	1786
HITACHI SR8000-F1/112(375MHz)	leibniz	112	1035.0	120000	15160	1344